# Cassandra Crossing/ Archivismi: upload ed operazioni elementari

(560 )—Continuiamo la nostra esplorazione di Archive.org per iniziare a capire come funziona e come usarlo.

---

**Cassandra Crossing/ Archivismi: upload ed operazioni elementari**
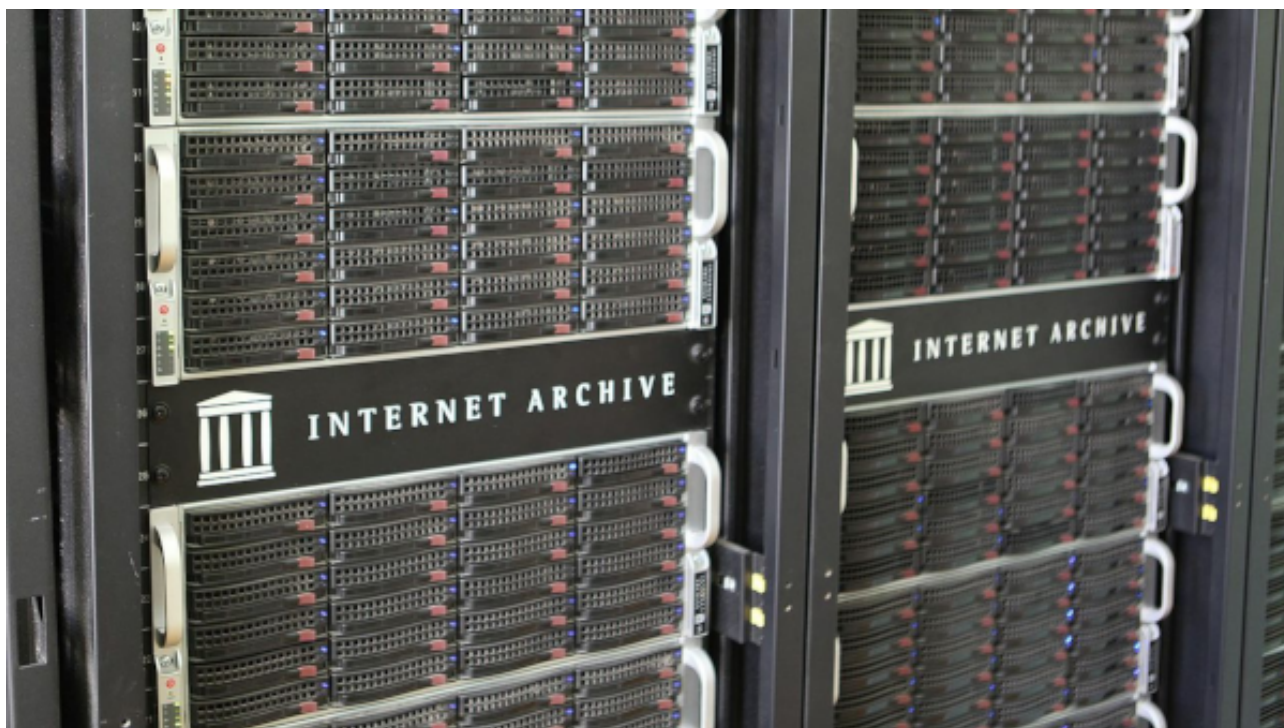
Figure 1:

*(560)—Continuiamo la nostra esplorazione di Archive.org per iniziare a capire come funziona e come usarlo.*

**26 dicembre 2023**— *Archiviare* non significa memorizzare. *Archiviare* non significa copiare.

*Archiviare*, nel mondo digitale e nel senso più esteso del termine, vuol dire memorizzare in forma significativa un'informazione digitale, e farlo nei formati più opportuni, corredandola dei metadati più adatti e da un ben selezionato insieme di parole chiave di ricerca. E fare questo seguendo le procedure ed i metodi consolidati che generazioni di bibliotecari, oggi *digitali* ma prima anche *analogici*, hanno già predisposto per noi.

In questa terza puntata di *Archivismi* scopriremo che archiviare su Archive.org non è semplice ed immediato come copiare un ebook od un video mp4 su Dropbox, Google Disk od un server Nextcloud.

Alcune operazioni sono, per fortuna, quasi completamente automatiche; come abbiamo visto nella scorsa puntata, archiviare una singola pagina su *The Wayback Machine* è in effetti un'operazione elementare, anche se un po' lenta. Ed in effetti è lenta perché sfrutta una infrastruttura complessa, archiviando la pagina con un meccanismo pensato per consentire anche operazioni molto più sofisticate.

Vediamo di cosa si tratta. Nel database di *Internet Archive* le informazioni sono memorizzate in oggetti. Ad ogni oggetto, al momento della creazione, viene associato un identificatore univoco. Un oggetto, a tutti gli effetti, può essere rappresentato come una directory, in cui sono contenuti almeno un file di dati ed almeno due file di metadati.

Proviamo a creare un oggetto eseguendo un semplice *upload*, come quelli che si utilizzano per caricare un file in un cloud.

Per proseguire, dovete aver creato il vostro utente di Internet Archive; se non lo avete già fatto, fatelo adesso e poi entrate col vostro utente.

Osservate subito il cuoricino  che sta al centro della barra dei menù; cliccandovi sopra potete effettuare una piccola donazione con qualsiasi mezzo di pagamento abbiate disponibile. Non è ovviamente obbligatorio, i servizi di Internet Archive sono gratuiti, come è giusto che sia in qualunque *biblioteca universale*, ma a loro far funzionare la baracca costa soldi, quindi, come al solito, TANSTAAFL.

Se invece per ora non vi sembra che il servizio di *Internet Archive* valga i vostri spiccioli, procedete pure; probabilmente presto cambierete idea.

Osservate in alto a destra il link UPLOAD; tra parentesi notate, e lo vedremo molte volte, che Internet Archive nasconde i link più importanti nei posti meno visibili, ma deve essere un'*arte oscura* comune tra i bibliotecari digitali ...

Se ci cliccate, si apre ovviamente una finestra in cui potete fare il drag&drop di un file od aprire una più pratica finestra di selezione file. Per seguire questo esempio, selezionate un file .pdf, oppure quello che volete voi.

Fatta la selezione, vi si aprirà la finestra più importante in assoluto, quella di *archiviazione*.

Figure 2:

Innanzitutto **non infestate Internet Archive con le vostre prove**; anche se è possibile cancellare un oggetto, in realtà questo non viene normalmente rimosso dal database, ma con-

tinua ad esistere per una serie di motivi. Potrebbe essere rimosso successivamente, ma questo richiede un intervento di tipo "sistemistico". Ci torneremo più avanti.

Per praticità è quindi previsto un apposito parametro, che marca un oggetto come *effimero*, e lo cancella automaticamente e completamente dopo 30 giorni. **Avete il dovere di fare i bravi**, quindi usate questo parametro per tutte le archiviazioni di prova, in modo che solo quelle definitive vengano memorizzate permanentemente.

Quindi selezionatelo subito dalla apposita tendina. Osserviamo poi che, ricavandoli dal nome del file, vengono proposti dei valori ragionevoli, come titolo dell'oggetto ed identificatore (URL).

Correggiamoli, se lo riteniamo necessario, ed aggiungiamo una descrizione, delle parole chiave, e quanto altro utile.

Tutti i campi sono facoltativi, ma farne un utilizzo attento e ragionato è il fattore essenziale in una buona campagna di archiviazione. E "*progettarla*" è anche la cosa più difficile, quindi per ora accontentiamoci. Ne riparleremo.

Potete ora cliccare sul pulsante azzurro in basso.

Figure 3:

Si apre una normale finestrina di upload che, dopo un tempo che vi parrà troppo lungo (minuti), vi restituirà una *vista* sull'oggetto che avete appena creato. Questa finestra include un browser per l'oggetto appena creato, ma ora vi apparirà probabilmente solo un riquadro bianco, perché in realtà il processo di archiviazione è ancora in corso.

Figure 4:

Esaminandolo con attenzione noterete tutta una serie di link cliccabili, ma prima una cosa importante.

A seconda di che browser e sistema operativo utilizziate, navigando indietro con la freccia a sinistra (si può fare tranquillamente) potrebbe accadervi che, oltre a visualizzare la pagina precedente, vi si apra la finestra di download del file; in questo caso potete annullarla/chiuderla tranquillamente e continuare. Quando Cassandra troverà un modo di evitare questo fastidio, certamente ve lo farà sapere.

**Il vostro oggetto di prova non è stato ancora completamente creato**; esiste come identificatore e come informazioni di base, e può perciò già essere utilizzato, ma molte operazioni nel backend di Internet Archive devono ancora essere eseguite, e lo saranno nei prossimi minuti, ore o giorni. Quindi, ancora una volta, pazienza.

Ma di quali operazioni si tratta? Dipende dal tipo di oggetto che avete creato, ed in quale "*collezione*" lo avete inserito. Tralasciamo per ora l'importantissimo aspetto della *collezione,* e concentriamoci sulle operazioni automatiche che sono state schedulate e che vengono o verranno compiute sull'oggetto appena creato. E' possibile esaminarle, utilizzando il link *history* nel microscopico menu in alto a sinistra nella finestra *oggetto*.

My Tasks Not Yet Completed for cassandra-crossing-2558-
di-cassandra-archivismi (page drawn: PST: 2023-12-26 03:31:25)

Legend and row counts:

| | |
|---|---|
| Waiting to run | 0 |
| Running | 1 |
| Waiting for admin | 0 |
| Parked | 0 |
| Other | 0 |
| Finished | 0 |
| server readonly -- tasks waiting for harddrive fix | |
| disk readonly -- tasks waiting for rescue task | |

where am I in line?

| identifier | task_id | server | cmd | submittime | submitter | args |
|---|---|---|---|---|---|---|
| cassandra-crossing-2558-il-dizionario-di-cassandra-archivismi - History Mgr | 4099976267 | iw601907 | derive.php | (17.3 minutes) | calamarim | server_primary=ia601204.u.. |

Figure 5:

C'è una task in running; si tratta dell'archiviazione dell'oggetto che procede, mentre nella parte bassa della finestra compare, e continua a popolarsi, lo storico delle operazioni eseguite automaticamente sull'oggetto; infatti dopo una mezz'ora apparirà questo.

Item History for cassandra-crossing-2558-il-dizionario-di-
archivismi (page drawn: PST: 2023-12-26 03:33:49)

Legend and row counts:

| | |
|---|---|
| Waiting to run | 0 |
| Running | 1 |
| Waiting for admin | 0 |
| Parked | 0 |
| Other | 0 |
| Finished | 3 |
| server readonly -- tasks waiting for harddrive fix | |
| disk readonly -- tasks waiting for rescue task | |

where am I in line?

| identifier | task_id | server | cmd | submittime | submitter | args |
|---|---|---|---|---|---|---|
| cassandra-crossing-2558-il-dizionario-di-cassandra-archivismi - History Mgr | 4099976267 | iw601907 | derive.php | (19.7 minutes) | calamarim | server_primary=ia601204.u.. |
| cassandra-crossing-2558-il-dizionario-di-cassandra-archivismi - History Mgr | 4099976257 | ia601204.us.archive.org | book_op.php | (19.7 minutes) | calamarim | op0=VirusCheck&dir=/24/it.. |
| cassandra-crossing-2558-il-dizionario-di-cassandra-archivismi - History Mgr | 4099973598 | ia601204.us.archive.org | archive.php | (22.6 minutes) | calamarim | done=delete&from_url=rsyn.. |
| cassandra-crossing-2558-il-dizionario-di-cassandra-archivismi - History Mgr | 4099973556 | ia601204.us.archive.org | archive.php | (22.7 minutes) | calamarim | done=delete&from_url=rsyn.. |

Figure 6:

Molte cose continueranno a succedere al nostro oggetto nel backend, e ne riparleremo; intanto torniamo nella finestra oggetto, e nel microscopico menù in alto a sinistra clicchiamo su *manage*.

Compariranno due grosse icone; quella di sinistra (importantissima) vi permette di editare i metadati, ma per ora è bloccata dal processo di creazione dell'oggetto in corso, e se ci cliccate vi verrà spiegato perché.

Quella di destra invece permette di editare i file contenuti nell'oggetto, e se la cliccate vi aprirà una vista sulla cartella e sul suo contenuto. A seconda del tempo passato e del file che avere archiviato, troverete contenuti diversi, e molti più file di quelli che vi aspettereste.
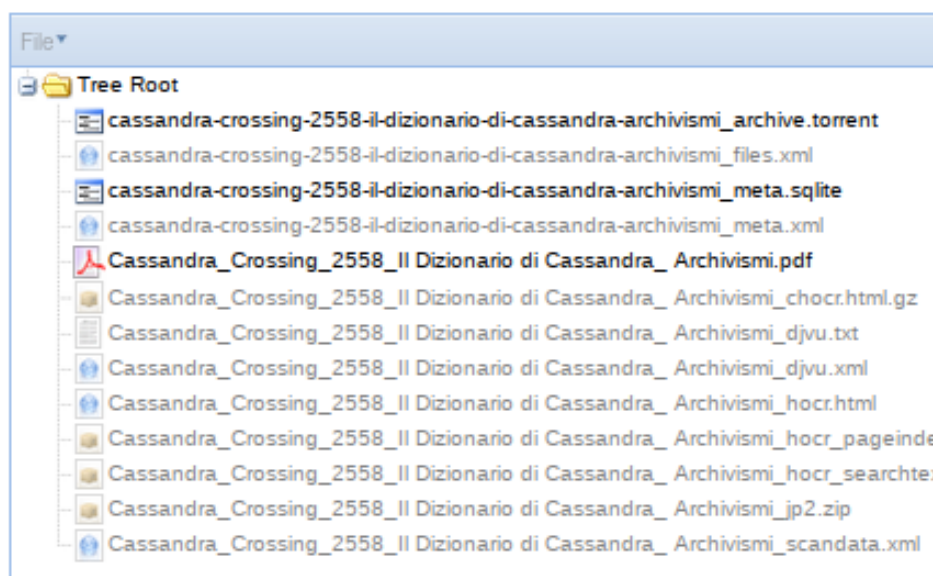
Figure 7:

Noterete il file .pdf originale che è stato usato in questo esempio, due file .xml ed uno .sqlite, che contengono informazioni di sistema (e come tante altre cose, ne riparleremo). C'è un nuovo file .torrent, che può già essere scaricato ed utilizzato per fornire un link *torrent*, utile se il file caricato fosse molto grande e lo si dovesse far scaricare da molte persone.

Ci sono infine diversi file, in parte ancora indicati in grigio ed inaccessibili, che *testimoniano* le operazioni ancora in corso che Internet Archive sta facendo per voi, e che dipendono dal tipo di file che avete archiviato.

Ad esempio, dal nostro file .pdf verrà creato automaticamente un file di solo testo, contenente appunto tutto il testo presente nel pdf. Sempre nel caso di un pdf verrà creato un indice della pagine. Se si fosse invece trattato di un file video, tra le altre cose sarebbe stata creata una directory contenente 255 thumbnail, uniformemente estratti da tutta la lunghezza del video, che possono essere usati per visualizzarlo come oggetto *video* (ad esempio in una *timeline*). Altri file verranno creati, ma ci avviciniamo alla fine di questa intensa puntata.

Perché … *questa è un'altra storia.*

Ma un'ultima cosa. Sempre dal peculiare *micromenù* in alto a sinistra della finestra *oggetto* si può accedere al link che apre la finestra dell'*item manager*, in cui è possibile gestire l'oggetto creato in molteplici aspetti.

Figure 8:

Alcuni tra i 24 indomiti lettori, i più interessati e dotati di tempo ed iniziativa, potranno partire da qui o dalle altre finestre che abbiamo visto per un'esplorazione *in solitario*, che potrà durare tantissimo e portarli molto lontano.

A questi arditi Cassandra raccomanda di dotarsi di un po' di Python e di confidenza con le API; a questo fine suggerisce di utilizzare la molto ben organizzata pagina di help e consegna loro questo prezioso link alla documentazione sviluppatori di Internet Archive.

Come esempio, se voleste sviscerare l'argomento dei file creati automaticamente durante un upload, potreste leggere questo articolo dell'help.

Gli altri aspetteranno invece che Cassandra, *lento pede*, compia questa esplorazione per loro od insieme a loro.

*Stay tuned* per la prossima puntata di "*Archivismi*".

---

Scrivere a Cassandra—Twitter—Mastodon
Videorubrica "Quattro chiacchiere con Cassandra"
Lo Slog (Static Blog) di Cassandra
L'archivio di Cassandra: scuola, formazione e pensiero



By Marco A. L. Calamari on December 26, 2023.

Canonical link

Exported from Medium on August 27, 2025.